JP2002163273

Publication Title:

DOCUMENT CONTROL METHOD AND SYSTEM

Abstract:

Abstract of JP2002163273

PROBLEM TO BE SOLVED: To solve a problem in browsing retrieved documents by using classified information wherein a browser takes time for determining them as a required document due to being unable to quickly understand why the classification is imparted to the document. SOLUTION: The correspondence between the classified information and a document recording place is informed to a user when displaying the retrieved result content by asking the user to designate the correspondence and holding the information with correspondence to the document.

Data supplied from the esp@cenet database - Worldwide

Courtesy of http://v3.espacenet.com

This Patent PDF Generated by Patent Fetcher(R), a service of Stroke of Color, Inc.

(19)日本国特許庁 (JP)

(12) 公開特許公報(A)

(11)特許出願公開番号 特開2002-163273 (P2002-163273A)

(43)公開日 平成14年6月7日(2002.6.7)

(51) Int.Cl. ⁷		識別記号	FΙ	FΙ		
G06F	17/30	2 1 0	C06F	17/30	210D	5B075
		170			1. 7 0 A	5B082
	12/00	5 2 0		12/00	520E	

審査請求 未請求 請求項の数26 OL (全 13 頁)

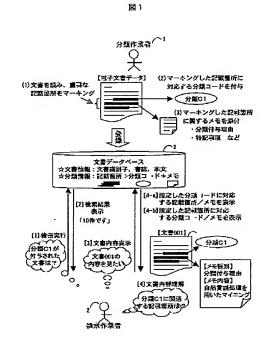
		音江明水	木明水 開水気の数20 〇L(主 13 頁)
(21)出願番号	特願2000-360710(P2000-360710)	(71)出願人	000005108
			株式会社日立製作所
(22) 出顧日	平成12年11月22日(2000.11.22)		東京都千代田区神田骏河台四丁目6番地
		(72)発明者	間瀬 久雄
			神奈川県川崎市麻生区王禅寺1099番地 株
			式会社日立製作所システム開発研究所内
		(72)発明者	久連石 一毅
			東京都江東区新砂一丁目6番27号 株式会
			社日立製作所公共システム事業部内
		(74)代理人	100075096
			弁理士 作田 康夫
			最終頁に続く

(54) 【発明の名称】 文書管理方法およびシステム

(57)【要約】

【課題】分類情報を用いて文書検索した結果を閲覧する際に、何故その分類がその文書に付与されたのかを迅速に理解できないため、利用者が必要とする文書であるか否かの判定に時間がかかる。

【解決手段】分類情報および文書記載箇所との対応を利用者に指定させ、その情報を文書と対応付けて保持し、 検索結果内容を表示する際にその対応関係を利用者に報知する。



【特許請求の範囲】

【請求項1】入力装置と、出力装置と、一つ以上の電子 文書データを含む文書データベースを格納する記憶装置 とを含む文書管理システムにおける文書管理方法であっ て、

前記入力装置を介して電子文書データの構成要素を指定 し、前記指定された構成要素に対応する分類情報を、前 記入力装置を介して入力し、前記指定された構成要素と 前記入力された分類情報とを対にして当該電子文書デー タと対応付けて前記記憶装置に格納することを特徴とす る文書管理方法。

【請求項2】入力装置と、出力装置と、一つ以上の電子 文書データを含む文書データベースを格納する記憶装置 と、前記入力装置を介して入力された検索条件に基づい て前記文書データベースから電子文書データを検索する 文書検索手段とを含む文書管理システムにおける文書管 理方法であって、

前記文書データベース中の個々の電子文書データについて、電子文書データの構成要素と当該構成要素に対応する分類情報の対を前記記憶装置に格納しておき、前記文書検索手段によって検索された前記電子文書データの内容を、前記出力装置を介して表示する際に、前記格納された電子文書データの構成要素と分類情報の対の中から当該電子文書データに該当する対を抽出し、前記抽出した構成要素と分類情報の対を対応付けて表示することを特徴とする文書管理方法。

【請求項3】入力装置と、出力装置と、一つ以上の電子 文書データを含む文書データベースを格納する記憶装置 と、前記入力装置を介して入力された検索条件に基づい て前記文書データベースから電子文書データを検索する 文書検索手段とを含む文書管理システムにおける文書管 理方法であって、

前記入力装置を介して電子文書データの構成要素を指定 し、前記指定された構成要素に対応する分類情報を、前 記入力装置を介して入力し、前記指定された構成要素と 前記入力された分類情報とを対にして当該電子文書デー タと対応付けて前記記憶装置に格納し、

前記文書検索手段によって検索された前記電子文書データの内容を、前記出力装置を介して表示する際に、前記格納された電子文書データの構成要素と分類情報の対の中から当該電子文書データに対応する対を抽出し、前記抽出した構成要素と分類情報の対を対応付けて表示することを特徴とする文書管理方法。

【請求項4】請求項1,請求項2あるいは請求項3記載の文書管理方法において、

前記出力装置を介して前記構成要素と分類情報の対を対応付けて表示する際に、前記入力装置を介して前記構成要素を指定すると当該構成要素に対応する分類情報を前記出力装置を介して表示し、前記入力装置を介して前記分類情報を指定すると当該分類情報に対応する構成要素

を前記出力装置を介して表示することを特徴とする文書 管理方法。

【請求項5】請求項1,請求項2あるいは請求項3記載の文書管理方法において、

前記入力装置を介して前記構成要素に関する文字情報を 入力可能とし、前記文字情報を前記構成要素と対応付け て前記記憶装置に格納し、前記出力装置を介して前記構 成要素を表示する際に前記対応する文字情報を表示する ことを特徴とする文書管理方法。

【請求項6】請求項1あるいは請求項3記載の文書管理 方法において、

前記指定された構成要素に対して前記入力装置を介して 属性を付与し、前記付与された属性を当該構成要素と対 応付けて前記記憶装置に格納することを特徴とする文書 管理方法。

【請求項7】請求項6記載の文書管理方法において、 前記入力装置を介して指定した属性値を持つ前記構成要 素を検索し、前記出力装置を介して前記検索結果を表示 することを特徴とする文書管理方法。

【請求項8】請求項6記載の文書管理方法において、 前記入力装置を介して指定した属性値を持つ前記構成要 素を収集し、前記収集した構成要素群から言語的あるい は構造的特徴を抽出し、前記出力装置を介して前記抽出 した特徴を表示することを特徴とする文書管理方法。

【請求項9】請求項1あるいは請求項3記載の文書管理 方法において、

電子文書データの構成要素の特徴を分類別に記述した分類知識を記憶装置に格納しておき、前記指定された構成要素の特徴を解析し、前記分類知識との照合により当該構成要素に対応する分類候補を認定し、前記出力装置を介して前記分類候補を報知することを特徴とする文書管理方法。

【請求項10】請求項9記載の文書管理方法において、前記報知された分類候補の中から正しい分類情報を前記入力装置を介して指定し、前記指定された分類情報と当該構成要素とを対にして当該電子文書データと対応付けて前記記憶装置に格納することを特徴とする文書管理方法

【請求項11】請求項9記載の文書管理方法において、前記報知された分類候補の中から正しい分類情報を前記入力装置を介して指定し、前記指定された分類情報と当該分類情報に対応する構成要素を解析して得られる特徴に基づいて前記分類知識の内容を更新すること特徴とする文書管理方法。

【請求項12】請求項2あるいは請求項3記載の文書管理方法において、

前記構成要素に対応して付与されている分類情報が誤り であると判断した場合、入力装置を介して正しい分類情報を入力し、前記入力された正しい分類情報に基づき前 記記憶装置に格納されている構成要素および分類情報の データを修正することを特徴とする文書管理方法。

【請求項13】請求項2あるいは請求項3記載の文書管理方法において、

前記文書検索手段によって検索された文書識別子集合を、前記出力装置を介して表示する際に、前記検索条件として使用された分類情報に対応する構成要素を当該文書識別子と対にして表示することを特徴とする文書管理方法。

【請求項14】入力装置と、出力装置と、一つ以上の電子文書データを含む文書データベースを格納する記憶装置とを含む文書管理システムにおいて、

前記入力装置を介して電子文書データの構成要素を指定する手段と,前記指定された構成要素に対応する分類情報を前記入力装置を介して入力する手段と,前記指定された構成要素と前記入力された分類情報とを対にして当該電子文書データと対応付けて前記記憶装置に格納する手段とを備えたことを特徴とする文書管理システム。

【請求項15】入力装置と、出力装置と、一つ以上の電子文書データを含む文書データベースを格納する記憶装置と、前記入力装置を介して入力された検索条件に基づいて前記文書データベースから電子文書データを検索する文書検索手段とを含む文書管理システムにおいて、

前記文書データベース中の個々の電子文書データについて、電子文書データの構成要素と当該構成要素に対応する分類情報の対を前記記憶装置に格納する手段と,前記文書検索手段によって検索された前記電子文書データの内容を、前記出力装置を介して表示する際に、前記格納された電子文書データの構成要素と分類情報の対の中から当該電子文書データに該当する対を抽出する手段と,前記抽出した構成要素と分類情報の対を対応付けて表示する手段とを備えたことを特徴とする文書管理システム。

【請求項16】入力装置と、出力装置と、一つ以上の電 子文書データを含む文書データベースを格納する記憶装 置と、前記入力装置を介して入力された検索条件に基づ いて前記文書データベースから電子文書データを検索す る文書検索手段とを含む文書管理システムにおいて、 前記入力装置を介して電子文書データの構成要素を指定 する手段と、前記指定された構成要素に対応する分類情 報を前記入力装置を介して入力する手段と、前記指定さ れた構成要素と前記入力された分類情報とを対にして当 該電子文書データと対応付けて前記記憶装置に格納する 手段と、前記文書検索手段によって検索された前記電子 文書データの内容を前記出力装置を介して表示する際 に、前記予め格納された電子文書データの構成要素と分 類情報の対の中から当該電子文書データに対応する対を 抽出する手段と、前記抽出した構成要素と分類情報の対 を対応付けて表示する手段とを備えたことを特徴とする 文書管理システム。

【請求項17】請求項14、請求項15あるいは請求項

16記載の文書管理システムにおいて、

前記出力装置を介して前記構成要素と前記分類情報を対応付けて表示する際に、前記入力装置を介して前記構成要素を指定する手段と,前記指定された構成要素に対応する分類情報を表示する手段と,前記入力装置を介して前記分類情報を指定する手段と,前記指定された分類情報に対応する構成要素を表示する手段とを備えたことを特徴とする文書管理システム。

【請求項18】請求項14,請求項15あるいは請求項 16記載の文書管理システムにおいて、

前記入力装置を介して前記構成要素に関する文字情報を 入力する手段と,前記文字情報を前記構成要素と対応付 けて前記記憶装置に格納する手段と,前記出力装置を介 して前記構成要素を表示する際に前記対応する文字情報 を表示する手段とを備えたことを特徴とする文書管理シ ステム

【請求項19】請求項14あるいは請求項16記載の文書管理システムにおいて、

選択された構成要素に対して前記入力装置を介して属性 を付与する手段と、前記付与された属性を当該構成要素 と対応付けて前記記憶装置に格納する手段とを備えたこ とを特徴とする文書管理システム。

【請求項20】請求項19記載の文書管理システムにおいて、

前記入力装置を介して指定した属性値を持つ前記構成要素を検索する手段と,前記検索結果を、前記出力装置を介して表示する手段とを備えたことを特徴とする文書管理システム

【請求項21】請求項19記載の文書管理システムにおいて

前記入力装置を介して指定した属性値を持つ前記構成要素を収集する手段と,前記収集した構成要素群から言語的あるいは構造的特徴を抽出する手段と,前記出力装置を介して前記抽出した特徴を表示する手段とを備えたことを特徴とする文書管理システム。

【請求項22】請求項14あるいは請求項16記載の文書管理システムにおいて、

電子文書データの構成要素の特徴を分類別に記述した分類知識を記憶装置に格納する手段と、前記指定された構成要素の特徴を解析する手段と、前記分類知識との照合により当該構成要素に対応する分類候補を認定する手段と、前記出力装置を介して前記分類候補を報知する手段とを備えたことを特徴とする文書管理システム。

【請求項23】請求項22記載の文書管理システムにおいて

前記報知された分類候補の中から正しい分類情報を、前 記入力装置を介して指定させる手段と、前記指定された 分類情報と当該構成要素とを対にして当該電子文書デー 夕と対応付けて前記記憶装置に格納する手段とを備えた ことを特徴とする文書管理システム。 【請求項24】請求項22記載の文書管理システムにお いて

前記報知された分類候補の中から正しい分類情報を前記 入力装置を介して指定させる手段と,前記指定された分 類情報と当該分類情報に対応する構成要素を解析する手 段と,前記解析手段により得られる言語的あるいは構造 的特徴に基づいて前記分類知識の内容を更新する手段と を備えたこと特徴とする文書管理システム。

【請求項25】請求項15あるいは請求項16記載の文書管理システムにおいて、

前記構成要素に対応して付与されている分類情報が誤り であると判断した場合に入力装置を介して正しい分類情報を入力する手段と,前記入力された正しい分類情報に 基づき前記記憶装置に格納されている構成要素および分類情報のデータを修正する手段とを備えたことを特徴と する文書管理システム。

【請求項26】請求項15あるいは請求項16記載の文 書管理システムにおいて、

前記文書検索手段によって検索された文書識別子集合を、前記出力装置を介して表示する際に、前記検索条件として使用された分類情報に対応する構成要素を当該文書識別子と対にして表示する手段を備えたことを特徴とする文書管理システム。

【発明の詳細な説明】

[0001]

【発明の属する技術分野】本発明は、新聞記事や論文,特許,Webページ等の電子文書データを管理する文書管理システムに係り、特に各々の電子文書データを分野や内容等によって分類された情報を検索条件に加味して検索する機能を持つ文書管理システムにおける文書分類/検索作業支援に関する。

[0002]

【従来の技術】分類情報を用いて文書検索を行う場合の 文書分類/検索作業支援に関する従来技術としては、文 書内容と文書属性との対応付け表示、文書検索結果内容 の表示、文書の自動分類、分類情報を用いた文書検索、 要約情報を用いた文書検索、という五つの技術がある。 【0003】(1)文書内容と文書属性との対応付け表

文書の構成要素をある文書属性と対応付けて表示する技術としては、文書編集システムにおいて、文書の章節項情報を抽出して本文とは別にリスト表示し、リスト中の項目を選択すると対応する章節項の先頭へジャンプできるという機能がある。また、Webページの記述言語として使われているHTML(Hyper TextMarkup Language)文書におけるハイパーリンクは、文書の構成要素(アンカー)に文書アドレス情報を対応付けたものである。

【0004】(2)文書検索結果内容の表示 検索された文書の内容を表示する技術としては、市販さ れている多くの文書検索システムが採用しているよう に、検索条件として使用されたキーワードを文書中でハ イライト表示する機能がある。また、キーワードが文書 全体のどのあたりに集中して出現しているかの分布デー タをビジュアルに表示する機能がある。

【0005】(3)文書の自動分類

文書管理システムの中には、文書の内容に応じて文書に分類コードを付与可能とするものがある。分類コードの付与方法としては、人手で付与するものと自動分類による付与がある。自動付与する技術としては、例えばSalt on, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley Publishing Company Inc. (1989)に示されているように、文書に出現するキーワードの出現傾向をベクトルで表し、分類知識として格納されているカテゴリ別のキーワードの出現傾向を表すベクトルとの類似度を計算することにより、付与すべき分類を推定する方式などの手法がある。

【0006】(4)分類情報を用いた文書検索 分類情報を用いた文書検索としては、http://www.yaho o.co.jp/等に代表されるように、予め分類体系を定義し ておき、各文書を関連する分類に対応付けておき、利用 者は分類体系を辿ったり、分類情報をキーとして文書を 絞り込み検索することにより、必要な文書を効率良く取 得するものがある。

【0007】(5)要約情報を用いた文書検索 検索された文書識別子集合を表示する際に、各文書の要 約を文書識別子とともに表示することにより、情報の取 捨選択を支援する文書検索システムがある。要約文章は 人手で作成するものと自動作成されるものがある。自動 作成方式としては、文書の冒頭部分を自動抽出したり、 文書の言語的特徴や構造的特徴に着目して自動生成した りするものがある。

[0008]

【発明が解決しようとする課題】本発明は、文書検索/ 分類機能を備えた上記従来の文書管理システムが抱える 以下の課題を解決する。

【0009】(a)文書分類時の作業ノウハウの活用 上記従来技術(3)および(4)で述べたような文書分 類機能を備えた従来の文書管理システムでは、分類コー ドのみを文書識別子に対応させて蓄積している。そのた め、文書中のどの記載箇所に着目して分類を付与したの かに関する情報や、何故その分類がその文書に付与され たのかなどに関する情報が得られない。また文書検索に おいて、分類情報を検索条件に加味して検索すること は、所望の文書を効率良く検索するのに有効であるが、 検索された文書内容を表示する際に、検索条件として用 いた分類情報に対応する記載が文書中のどこにあるのか がすぐに分からないので、文書をある程度読まないとそ の文書が所望の文書であるか否かを判断できない。この ように、分類作業者の意図が検索作業者に十分に伝わらないことが、検索作業者の検索作業効率および質の向上を妨げている。

【0010】また、上記従来技術(2)で述べたような 文書検索結果表示機能は、文書の取捨選択には有効であ るが、分類コードそのものは文書本文中に明示されない ため、この方式では検索条件として使用された分類に対 応する文書中の記載箇所を特定することができない。

【0011】さらに、上記従来技術(1)で述べたハイパーリンクは、リンク方向が一方通行なので、分類情報から文書記載部分,文書記載部分から分類情報へという双方向的な対応付け表示ができない。また、上記従来技術(5)で述べた要約情報を付加した従来の文書検索システムは、要約内容が分類情報と対応付けられていないので、分類に関連する記載箇所を特定する手掛かりにはならない。

【0012】(b)分類付与作業の低コスト化上記従来技術(3)で述べた従来の文書自動分類方式の多くは文書中のキーワードの出現傾向に着目しているが、文書全体を入力としているため、文書の主題を表していない部分に出現するノイズ語句が分類結果に与える影響は大きく、その結果、分類結果にノイズが多く含まれてしまう。この傾向は分類体系の粒度が細かくなるほど、また分類数が多くなるほど、顕著になる。従って、文書を分類する際には、分類で着目すべき文書記載箇所を限定し、個々の記載箇所について、対応する分類を付与することを支援する機能が必要となる。

【0013】また、分類付与作業において自動分類技術を採用する際には、自動分類に不可欠な分類知識(ルール)を定期的に保守/更新していく必要がある。従って、自動分類精度を高レベルで維持するのに十分な質を持った分類知識をいかに効率良く作成するかが課題である。

[0014]

【課題を解決するための手段】本発明では、以下の手段により、上記課題を解決する。利用者が電子文書データの構成要素を指定する手段と、指定された構成要素に対応する分類情報を利用者が入力する手段とを設け、指定された構成要素と入力された分類情報とを対にして電子文書データと対応付けて記憶装置に格納し、文書検索手段によって検索された電子文書データの内容を表示する際に、電子文書データの構成要素と分類情報の対の中から当該電子文書データに対応する対を抽出し、これらを対応付けて表示することにより、上記課題(a)で述べた課題を解決する。

【 0 0 1 5 】また、電子文書データの構成要素の特徴を 分類別に記述した分類知識を記憶装置に格納しておき、 利用者によって指定された構成要素の特徴を解析し、分 類知識との照合により当該構成要素に最も関連する分類 候補を認定し、利用者に報知することにより、上記課題 (b)で述べた分類付与作業の効率化の課題を解決する

【0016】さらに、利用者が対応付けた構成要素と分類情報の対から当該構成要素の特徴を解析し、解析結果に基づいて分類知識の内容を更新することにより、上記課題(b)で述べた分類知識の品質維持の課題を解決する。

[0017]

【発明の実施の形態】本発明の実施例について、以下図面を用いて詳細に説明する。なお、これにより本発明が限定されるものではない。

【0018】本実施例では大量の特許明細書データを対象とした文書管理システムについて述べるが、本発明は特許明細書データに限定されるものではなく、新聞記事や科学技術論文、Webページなどさまざまな種類の文書データの管理に適用できる。

【0019】図1は本実施例の概要を示す図である。本システムは、管理対象となる特許明細書(以下、文書と呼ぶ)について、分類作業者1が各文書に関連する分類情報を付与し、文書データベース3に登録することを支援する。また、検索作業者2が文書データベース3の中から所望の文書データを検索することを支援する。ここで、分類作業者1と検索作業者2は同一人物であっても別人物であってもどちらでも良いし、それぞれが複数人数いても構わない。

【0020】本実施例における「分類情報」とは、特定の情報が文書中のどこに記載されているかを示す「記載箇所」と、当該記載箇所に対応する「分類コード」、当該記載箇所あるいは当該分類コードに関する「メモ情報」が互いに関連付けられて構成されたデータを指す。もちろん、メモ情報が格納されていない分類情報や、分類コードが付与されていない分類情報も許容される。

【0021】本実施例では、図1に示すように以下の手順によって電子文書を文書データベース3に登録し、登録した文書を検索し、閲覧する。

【0022】まず、以下の手順で文書を文書データベース3に登録する。

- (1)文書読解:分類作業者1は出力装置を介して表示 された登録対象文書の内容を読む。そのとき、特筆すべ き記載箇所にマーク(印)を付ける。
- (2)分類コード付与:必要に応じて、マーキングした 記載箇所に関連する分類コードを記載箇所に対応させて 付与する。
- (3)メモ添付:必要に応じて、マーキングした箇所あるいは付与した分類コードに関するメモ情報を記述し、記載箇所および分類コードと対応させて格納する。

【0023】メモ情報の内容としては、その分類を付与した根拠/理由、その記載箇所に対する解釈、その他伝達事項、などが挙げられるが、特に限定されるものではない。

【0024】文書データベース3は、上記3種類からなる分類情報と、文書識別子、書誌情報、本文内容からなる文書情報とから構成される。もちろん、分類情報は文書情報と物理的に別の場所に格納されても良い。ただし、この場合、文書識別子等を介して両データを対応付けられるようにしておく必要がある。

【0025】文書検索および文書閲覧は以下の手順で行われる。

[1]検索実行:検索作業者2は、キーワードおよび分類コードを用いて検索条件を作成し、検索を実行する。

[2] 検索結果表示:検索実行結果として、検索結果件数と検索された文書識別子一覧を表示する。

[3] 文書内容表示: 検索作業者2は、検索結果の中から内容を閲覧したい文書識別子を選択し、内容表示させる。

[4]文書内容理解:表示された文書に対して、検索作業者2が検索条件の中で用いた分類コードあるいはその文書に付与されている分類コードの中から特定の分類コードを指定すると、その分類コードに対応する文書中の記載箇所を他と異なる態様で表示する。また、その記載箇所に対応するメモ情報があれば併せて表示する。逆に、検索作業者2がマーキングされている記載箇所の中から特定の記載箇所を指定すると、その記載箇所に対応する分類コードやメモ情報を表示する。

【0026】このように、検索文書を閲覧する際に、本文全体をそのまま表示するだけでなく、ある特定の分類 コードに対応した記載箇所を提示可能とすることによ

り、検索作業者2が必要とする文書であるか否かの取捨 選択を短時間で行える。また、メモ情報を参照すること により、重要情報の見落しを防ぐとともに、他の作業者 との間で分類や検索に関するノウハウを共有できる。

【0027】図2は、本実施例で述べる文書管理システムのブロック構成を示す図である。分類作業者1あるいは検索作業者2(以下、作業者と総称)は入出力部4を介してシステムとやりとりを行う。分類情報登録部5は、作業者が画面上に入力した分類情報(記載箇所、分類コード、メモ情報)を取得し、これらを対応付けて文書データベース3に格納する。取得したデータは一旦ワーキングエリア41(後述の図3参照)に一時的に格納され、分類情報が確定した時点で文書データベース3に格納される。

【0028】分類コード推定部6は、分類作業者1から 指定された記載箇所に関連する分類コード候補を分類知 識7を参照することにより推定し、結果を作業者に提示 する。分類知識7は、分類コード別に言語的あるいは構 造的特徴を格納したデータである。ここで、言語的特徴 とは、キーワードの出現傾向などを指し、構造的特徴と はその記載がどの章節項/タグに出現するかなどを指 す。分類知識更新部8は、分類情報登録部5で取得,登 録された分類情報を分類コード別に解析して言語的ある いは構造的特徴を抽出し、その結果に基づいて分類知識 7を更新する。

【0029】文書検索部9は、検索作業者2によって指定された検索条件に合致する文書データベース3中の文書を検索して結果を報知する。検索条件は、キーワードあるいは分類コードの論理的組合せ(AND/OR/NOT等)で構成される。この種の文書検索は非常に多くのシステムが市販されており、容易に実現可能であるので、ここでは深く言及しない。

【0030】分類情報検索部10は、作業者から指定された分類情報の構成要素(記載箇所,分類コード,メモ情報のどれか)に対応する他の構成要素を文書データベース3の分類情報から検索して結果を表示する。例えば、分類情報が登録されているある記載箇所が指定された時に、その記載箇所に対応する分類コードおよびメモ情報を文書データベース3から検索する。

【0031】分類情報修正部11は、文書に付与される 分類コードが確定された後に分類コードを修正したい時 に、文書データベース3中の分類コードデータを書き替 える。

【0032】図3は、本実施例で述べる文書管理システ ムのハードウェア構成を示す図である。本システムは、 プログラムおよびデータを格納する記憶装置40と、記 憶装置40からプログラムおよびデータをロードして演 算処理を実行する処理装置20,入力装置であるキーボ ード31とマウス32,出力装置である出力モニタ33 からなる。また、計算機ネットワーク34を介して本シ ステムを稼働させることも可能である。記憶装置40は さらに、ワーキングエリア41,分類情報登録部格納エ リア42,分類コード推定部格納エリア43,分類知識 更新部格納エリア44,文書検索部格納エリア45,分 類情報検索部格納エリア46、分類情報修正部格納エリ ア47、分類知識格納エリア48、文書データベース格 納エリア49からなる。このうち、ワーキングエリア4 1,分類知識格納エリア48および文書データベース格 納エリア49はデータを格納するエリアであり、その他 の格納エリアはプログラムを格納するエリアである。

【0033】図4は、本実施例で述べる文書管理システムにおける分類付与作業時の画面インタフェースの一例を示す図である。本画面50は、文書内容を表示するサブ画面51と表示された文書に付与された分類コードを表示するサブ画面64とからなる。

【0034】分類作業者1は、読込ボタン70を押下して作業対象の文書をロードする。ロードされた文書内容はサブ画面51に表示される。サブ画面51は、文書識別子を表示するエリア52、文書内容を表示するエリア58、選択された記載箇所に付与された分類コードを表示するエリア56、選択された記載箇所に添付されたメモ情報を表示するエリア57、選択された記載箇所以前にマークされた記載箇所にジャンプする前マークボタン

53、選択された記載箇所以降にマークされた記載箇所 にジャンプする次マークボタン54、表示される分類情 報の表示/非表示状態を切替えるボタン55から構成さ れる。分類作業者1は表示された内容を読み、特筆すべ き記載箇所をマウスでドラッグすることにより範囲指定 する。次に、マウス右ボタンをクリックすると、範囲指 定した記載箇所をマーキングするためのメニュー60, 61を表示する。マーキング60を選択すると、範囲指 定した部分を特筆すべき記載箇所としてワーキングエリ ア41に一時的に登録する。登録された記載箇所は記載 部分をマウスでクリックすることにより選択状態にする ことができる。マーキング61を選択すると、マーキン グされた記載範囲に対応する分類コードおよびメモ情報 を記述する画面(図5)が表示されるので、データを入 力することにより、記載箇所に対応する分類コードおよ びメモ情報をワーキングエリア41に一時的に登録す る。マーキング修正62は、既にマーキングされている 記載箇所のマーキング範囲を修正する。この際に、記載 箇所に対応して登録されている分類コードおよびメモ情 報はそのまま残される。マーキング解除63は、既にマ ーキングされている記載箇所のマーキングを解除する。 この際に、記載箇所に対応して登録されている分類コー ドおよびメモ情報は削除される。

【0035】分類作業者1によって登録された分類情報はサブ画面51に表示される。登録されている特定の記載箇所をマウスで選択すると、それに対応する分類コードがエリア56に表示され、メモ情報がエリア57に表示される。また、エリア56に表示される分類コードの中から一つをマウスでダブルクリックすると、当該分類コードが付与されている記載箇所を他と異なる態様でエリア58に表示する。

【0036】サブ画面64は、サブ画面51に表示されている文書に付与される分類コード一覧を表示する画面である。記載箇所に対応させずに分類コードを付与したい時には、入力エリア65に直接分類コードを入力し、追加ボタン66を押下することにより、分類コードがワーキングエリア41に一時的に登録され、その結果がエリア69に表示される。エリア69に表示された分類コードをマウスで選択して削除ボタン68を押下することにより削除することができる。また、エリア69に表示された分類コードの中から一つを指定して記載箇所表示ボタン68を押下することにより、当該分類コードを含む記載箇所すべてをエリア58に表示する。

【0037】保存ボタン71を押下することにより、それまでに入力され、ワーキングエリア41に一時保存されていた分類情報を文書データベース3に登録する。分類付与作業を終了する際には終了ボタン72を押下する。

【0038】図5は、本実施例で述べる文書管理システムにおける分類コード付与およびメモ入力時の画面イン

タフェースの一例を示す図である。上述したように、本 画面80は本文中の記載箇所を指定した後に表示され る。

【0039】付与すべき分類コードが既知である場合、分類作業者1は入力エリア82に分類コードを入力し、追加ボタン83を押下することにより、エリア86に追加結果が表示される。エリア86において特定の分類コードを選択し、削除ボタン85を押下すると、当該分類コードが削除される。

【0040】付与すべき分類コードが未知である場合、 分類作業者1は付与すべき分類コードを決定する処理を 行う。分類コード候補提示ボタン84を押下すると、分 類コード推定部6において記載箇所の内容を解析して言 語的特徴(キーワード、出現頻度等)および構造的特徴 (特定の文書タグに記載されている等)を抽出し、分類 知識7との照合により記載箇所に関連する分類コードを 推定し、分類候補リストとしてサブ画面92に表示す る。エリア96に表示された分類コードのうち、X印の 付いた分類コードがシステムによって推定された分類コ ードである。本実施例では、システムによって推定され た分類コードそのものだけでなく、その分類コードの上 位分類に相当する分類コードおよび兄弟分類(同一の上 位分類を持つもの)に相当する分類コードも併せて表示 することにより、正しい分類コードの選択効率を向上さ せる。分類作業者1は、システムが提示した分類コード 候補の中から正しいものを選択して、追加ボタン95を 押下すると、その時点で選択されたすべての分類コード がエリア86に表示される。特定の分類コードをマウス で選択して定義参照ボタン93を押下すると、その分類 コードの範囲を規定する文章を表示する。また、全表示 ボタン94を押すと、対象とする分類体系の一覧をエリ ア96に表示する。

【0041】一方、メモ入力については、エリア89およびエリア91にメモ情報を入力する。エリア89では当該記載箇所の種別を選択する。この情報は後述するように特定の種別を持つ記載箇所を後に検索/収集する際のキーとして使われる。エリア91にはメモの内容を記述する。メモ内容は任意のフォーマットで記述できるので、文章を記述しても情報のアドレス等の情報を記載しても良い

【0042】一通りの情報を入力したら、追加ボタン98を押下することにより、入力されたデータが記載箇所と対応付けられた形でワーキングエリア41に一時的に保存される。取消ボタン99を押下した場合は、入力されたデータは保存されない。

【0043】図6は、本実施例で述べる文書管理システムにおける検索作業時の画面インタフェースの一例を示す図である。本画面100は、検索条件を指定するサブ画面111,検索結果文書識別子集合を表示するサブ画面116,指定した文書内容を表示・閲覧するサブ画面

101から構成される。

【0044】検索作業者2は、まずサブ画面111で検索条件を入力する。本実施例では、検索条件はキーワードおよび分類コードの論理的組合せ(AND/OR/NOT)として記述される。例えば、「分類コード5B075CE01が付与され、かつキーワード"データ分析"を含む文書」を検索することができる。キーワードに関する検索条件をエリア112に入力し、分類コードに関する検索条件をエリア113に入力する。本実施例では両者の検索条件をともに満たす文書を検索する。図6では、検索条件として、分類コード5B075CE01と5B075EB04の両方を含む文書を検索するように設定されている。検索実行ボタン114を押下すると、検索条件に合致する文書を文書データベース3から検索し、検索結果文書集合をエリア120に出力する。またこのとき、検索条件に使われた分類コード情報をエリア117に表示する。

【0045】検索作業者2は、エリア120に表示され た文書識別子の一つを選択し、表示ボタン119を押下 すると、サブ画面101に文書内容を表示する。この 時、マーキングされたすべての記載箇所が他と異なる態 様で表示される。検索作業者2は、検索条件で使われた 分類コードに対応する記載箇所を特定するために、ま ず、エリア117に表示された分類コードを一つ(複数 でも良い)選択し、記載箇所表示ボタン118を押下す ると、選択された分類コードが付与された記載箇所のみ を他と異なる態様で全て表示する。また、マーキングさ れた記載箇所を一つ選択すると、その記載箇所に対応す る分類コードをエリア106に、メモ情報をエリア10 7に表示する。さらに、エリア106に表示された分類 コードの中から一つをマウスでダブルクリック選択する と、当該分類コードが付与されている記載箇所のみを他 と異なる態様でエリア108に表示する。

【0046】検索作業者2は記載箇所に対して付与され ている分類コードが誤っていることを発見した時、その 記載箇所をマウスで選択指定し、マウス右ボタンをクリ ックすることにより、分類コードを修正するためのメニ ュー110を表示する。分類コード修正110を選択す ると、図11に示すような分類コードを追加・削除する 画面が表示され、図5で説明したのと同様の手順により 分類コードを修正する。図11では特定の記載箇所に付 与された分類コードの修正のほか、文書全体に付与され ている分類コードの修正も行える。ただし、検索作業者 2の誰でもが任意の文書の任意の記載箇所に対応する分 類コードを自由に修正することを許容すると、データの 質の維持ができなくなる可能性があるので、その場合に は、検索作業者2のログインIDを参照してその作業者 が修正権限を持ったユーザであるか否かをチェックする などのユーザ認証機能を付ければ良い。

【0047】図7は、本実施例における文書管理システムの文書データベース3の分類情報の構成の一例を示す

図である。前述したように、文書データベースは文書情報(書誌情報,本文内容等)と分類情報(記載箇所,分類コード,メモ情報)とから構成される。これら2種類のデータは文書識別子によってリンク付けされる。

【0048】分類情報は、文書識別子301,記載箇所の内容を分類する記載種別302,記載開始位置303,記載終了位置304,記載箇所が図面だったときに図面が記載されているポインタ情報を示す図面記載位置305、記載箇所に付与された分類コード(上位分類)306,記載箇所に付与された分類コード(下位分類)307,メモ情報308からなり、メモ情報はさらにメモ作成者309とメモ内容310より構成される。

【0049】文書識別子301は、文書を一意に特定するキーとなる。記載種別302は、記載箇所をマーキングする時の根拠(視点)に相当する識別情報であり、例えば分類を付与する根拠となる記載であったり、これまでにない新しい技術についての記載であったり、後日もう一度参照したいという付箋であったりする。これらの識別情報は予めシステム管理者によって定義登録することは容易に可能である。

【0050】記載開始位置303および記載終了位置304は、記載箇所がテキスト情報である場合に、記載箇所の範囲を規定するデータであり、本実施例では文書の先頭からのバイト数を記述する。記載箇所がテキスト情報ではなく、イメージ情報のような場合は、記載開始/終了位置を特定できない。そこで、このような場合、図面記載位置305として、例えばイメージ情報が別ファイルで保存されている場合は、ファイル名を格納し、イメージ情報が座標で示される場合には、図面左上の座標の値を格納する。

【0051】上位分類306および下位分類307は、記載箇所に対応して付与された分類コードである。一つの記載箇所に対して複数種類の分類コードを対応付けても良い。また、一つの文書に対して複数の分類体系を用意し、全く独立に分類コードを付与しても良い。

【0052】メモ情報308は、記載箇所に対応するメモ情報を格納する。作成者309の情報は、本システムログイン時の情報を保持することによって、自動的に取得できる。また、メモ内容310は、テキストデータとして格納される。

【0053】なお、後述する分類情報検索部10などにおける処理において、文書データベース3を何度も参照する際には、文書データベース3から処理対象となる文書の情報をワーキングエリア41に文書ロード時に文書データと一緒に吸い上げておき、処理プログラムはワーキングエリア41にアクセスして必要なデータを参照することにより、処理性能を向上できる。

【0054】以下では、図2に示すブロック図を構成する個々の処理内容および処理アルゴリズムについて詳しく述べる。

【0055】まず、分類情報登録部5の処理アルゴリズムについて述べる。文書データベース3に分類情報を登録するのは、図4において保存ボタン71が押下された時である。それまでに分類作業者によって入力された分類情報はワーキングエリア41に一時的に格納されている。分類情報登録部5では、分類情報を取得してワーキングエリア41上に格納されているデータを更新する処理と、ワーキングエリア41に格納されているデータを文書データベース3に登録する処理を行う。

【0056】分類情報登録部5が起動されるのは主に以下のイベントが発生した時である。

(1)読込ボタン70が押下された。

読み込む文書に関する分類情報が文書データベース3に 存在するか否かを判別し、存在する場合、そのデータを ワーキングエリア41にロードする。

【0057】(2)マーキングメニュー60が押下された。

マーキング対象の文字列を探索し、文書先頭から文字列 先頭までのバイト数をカウントし、その値をワーキング エリア41に格納された分類情報の記載開始位置に格納 追加する。また、文書先頭から文字列末尾までのバイト 数をカウントし、その値をワーキングエリア41に格納 された分類情報の記載終了位置に格納追加する。

【0058】(3)マーキング修正メニュー62が押下された。

マーキング範囲修正の対象となる記載箇所について、範囲修正後の文字列を探索し、文書先頭から文字列先頭までのバイト数をカウントし、その値をワーキングエリア41に格納された分類情報の当該記載箇所の記載開始位置に上書き格納する。また、文書先頭から文字列末尾までのバイト数をカウントし、その値をワーキングエリア41に格納された分類情報の当該記載箇所の記載終了位置に上書き格納する。

【0059】(4)マーキング解除メニュー63が押下された。

解除対象となる記載箇所に関するワーキングエリア41 中のデータを削除する。

【0060】(5)分類コード追加ボタン66,83,95、98が押下された。

それぞれ特定エリアに記述された、追加対象の分類コード情報を取得し、ワーキングエリア41中の当該記載箇所に対応する分類コードとして追加格納する。

【0061】(6)分類コード削除ボタン68,85が押下された。

それぞれ特定エリアで選択された、削除対象の分類コード情報を取得し、ワーキングエリア41中の当該記載箇所に対応する分類コードを削除する。

【0062】(7)保存ボタン71が押下された。 保存ボタン71が押下された時点でワーキングエリア4 1に格納されている分類情報をすべて文書データベース 3に格納する。既に登録されている分類情報に関するデータについては上書きで格納し、新規の分類情報の場合は文書識別子とともに新しいレコードに格納する。

【0063】図8は、分類知識7の構成の一例を示す図 である。分類知識7は各分類コードの言語的特徴あるい は構造的特徴を記述したデータである。言語的特徴とし てはキーワードや特定の言い回しの出現傾向が、構造的 特徴としては記載箇所の文書中の出現位置が挙げられ る。図8では、マーキングされた記載箇所から抽出され たキーワード701が分類コード702別の重みの値7 03とともに記述されている。重みの値703は、例え ばある特定のキーワードがマーキングされた記載箇所に 出現する回数に占める、ある特定の分類コードが付与さ れてマーキングされた記載箇所に出現する回数の割合を もって算出することができる。ある特定の分類コードに 関する記載箇所にしか出現しないキーワードは分類コー ドを特定するのに有効な情報となり、逆に多くの分類コ ードにまたがって出現するキーワードは比較的重要でな いキーワードとみなされる。

【0064】分類知識データのすべてを、マーキングされた記載箇所を解析することにより抽出するのは、マーキングされた記載箇所が少ない段階では抽出されるキーワード数が少なく分類精度が悪くなるので、初期の分類知識データは、文書全体について付与されている分類コード情報を用いて生成し、マーキングされた記載箇所の量が十分に大きくなった時点で記載箇所の解析による質の高い分類知識データに置き換えるという方式でも良い

【0065】図9は、分類コード推定部6の処理アルゴリズムを示す図である。まず、分類コード候補を推定する対象となる記載箇所文字列を取得する(ステップ601)。次に、文字列を形態素解析して文字列中のキーワードを抽出する(ステップ602)。ここではキーワードとして名詞すべてを抽出している。次に、キーワードの各々に重み1を付与する(ステップ603)。ここでは重みの値を一律1としているが、Tf/Idf法に代表されるようなキーワードの出現傾向から重みの値を算出する方式を採用しても良い。

【0066】次に、ステップ605の処理が済んでいないキーワードが残っているか否かを判別し(ステップ604)、残っている場合は、分類知識7を参照して当該キーワードが出現する分類コードを探索し、見つかった分類コードに対して当該キーワードの持つ重みの値を付与する(ステップ605)。残っていない場合は、ステップ606に進み、分類コード毎に付与された重みを合計することにより、その分類コードの得点を算出する(ステップ606)。最後に、得点の高い上位N件を分類コード候補として認定する(ステップ607)。自然数Nの値はシステム管理者によってチューニング可能である。

【0067】図10は、分類知識更新部8の処理アルゴリズムを示す図である。分類知識7の更新を、新しい分類情報が文書データベース3に登録された時点で行う場合と、ある一定の期間毎にまとめて更新する場合があるが、本実施例ではどちらでも構わない。前者は新しい情報が即座に分類知識7に反映されるが、分類知識更新処理の高速化が要求され、後者はその逆である。

【0068】分類知識7の更新は例えば以下の手順で実 現される。まず、分類知識7の更新に使われる記載箇所 の文字列および分類コードの対を文書データベース3か らすべて取得する(ステップ801)。次に、取得した 記載箇所の文字列をすべて形態素解析し、キーワード (名詞)を抽出する(ステップ802)。次に、抽出さ れたキーワードの重みを一律1に設定する(ステップ8 03)。次に、分類コード別にキーワードを纏め上げ る。このとき、同一キーワードの重みを合計する(ステ ップ804)。次に、キーワードと分類コード、重みか らなるリストをキーワードでソートし、あるキーワード がある分類コードを持つ記載箇所に出現する相対的割合 を算出し、その値をその分類コードにおけるそのキーワ ードの重みとして認定する(ステップ805)。こうし てできるキーワードと分類コードと重みの三つからなる 組を分類知識7とし、古い分類知識7と置き換える(ス テップ806)。

【0069】上記の分類知識7の更新処理では形態素解 析に時間がかかるので、形態素解析した結果を分類知識 生成後も保持し、次回の更新処理でその結果をそのまま 使うことにより、分類知識更新処理時間を短縮できる。 【0070】次に、分類情報検索部10の処理アルゴリ ズムについて述べる。分類情報検索部10は、図4,図 5,図6において、作業者から分類情報を構成する特定 のデータをトリガにしてそれに対応する別の分類情報構 成要素を文書データベース3から検索する。すなわち、 記載箇所をトリガとして、対応する分類コード、メモ情 報を検索したり、逆に、分類コードをトリガとして、対 応する記載箇所を特定する。ただし、前述したように、 図4、図5、図6では検索の対象はある特定の文書に関 する分類情報に限定されているので、読込ボタン70の 押下によって文書をロードする際に、文書データベース 3からその文書に関連する分類情報のみをワークエリア 41にコピーして保持することにより、データへのアク セス時間を短くできる。

【0071】分類情報検索部10が起動されるのは主に以下のイベントが発生した時である。

(1) エリア58, 108において特定の記載箇所が選択された。

選択された記載箇所の記載位置を特定し、ワークエリア 41にコピーされた文書データベース3の分類情報から 当該記載箇所に対応するレコードを特定し、対応する分 類コードおよびメモ情報を取得し、エリア56...57に 分類コードを、エリア57,107にメモ情報をそれぞれ表示する。

【0072】(2)エリア56,106の分類コードが ダブルクリックされた。

ワークエリア41にコピーされた文書データベース3の 分類情報の中から指定された分類コードが付与されてい るものを抽出し、対応する記載箇所の記載開始位置およ び終了位置を取得し、エリア58,108において記載 箇所を他と異なる態様で表示する。

【0073】(3)記載箇所表示ボタン67,118が押下された。

上記(2)と同様の処理手順により、エリア58,10 8において記載箇所を他と異なる態様で表示する。

【0074】(4)前マーク53,103,次マーク54,104が押下された。

前マーク53,103が押下された時には、押下される 直前に選択されていた記載箇所の記載開始位置よりも前 の記載開始位置である記載箇所の記載開始位置および記 載終了位置を、ワークエリア41にコピーされた文書デ ータベース3の分類情報の中から検索し、その記載箇所 を他と異なる態様で表示する。

【0075】次マーク54,104が押下された時には、押下される直前に選択されていた記載箇所の記載開始位置よりも後ろの記載開始位置である記載箇所の記載開始位置および記載終了位置を、ワークエリア41にコピーされた文書データベース3の分類情報の中から検索し、その記載箇所を他と異なる態様で表示する。

【0076】次に、分類情報修正部11の処理アルゴリズムについて述べる。分類情報修正部11が起動されるのは、図6において分類コード修正メニュー110が選択された時である。分類コード修正メニュー110を選択すると、図11に示す分類コードを追加・削除する画面が表示されるので、図5で述べたのと同様の方法で分類コードを修正する。図11で修正ボタン1111を押下すると、修正後の分類コードを図11の画面から取得し、文書データベース3の対応する分類コードを上書きして修正する。修正する際にユーザ認証が必要な場合は、ユーザ認証ファイルをアクセスして文書データベース3の修正が許可されているユーザであるかをチェックした後に、上記修正処理を行う。

【0077】次に、本実施例の拡張例を示す。

(1)記載箇所と分類コードとの対応情報を検索文書の 要約として利用

従来技術(5)で述べたように、各文書の要約を文書識別子とともに表示することにより、情報の取捨選択を支援する文書検索システムがあるが、ここでは、検索条件として使われた分類コードが付与されている記載箇所を要約として表示することにより、検索作業者は文書内容を表示して閲覧することなく、その情報が必要であるか否かを判定できる。文書検索部9によって出力される文

書集合の各々について、検索条件として使われた分類コードが付与されている記載箇所を、文書データベース3を参照することにより特定し、当該文書識別子と対にして図6のエリア120に表示する。複数の記載箇所がヒットした場合にすべての記載箇所を表示するか、先頭の記載箇所のみを表示する(さらに次々に記載箇所を表示するボタンを設ける)かを作業者に選択させることも容易に実現可能である。

【0078】(2)記載箇所に属性を指定可能とすることにより、特定情報を収集、分析

図5のエリア87に示すように、記載箇所に対して記載 内容の種別を選択定義できるようにしておくことによ り、ある特定の観点に立った記載箇所を大量の文書から 横断的に収集することが可能となる。例えば、特許明細 書を対象にした場合、これまでに見たことのないような 発明について記述された記載箇所に対して「新技術」と いう種別を付与しておく。「新技術」という種別が付与 された文書(記載箇所)を後に容易に検索できるように することにより、特許の出願傾向を迅速に分析すること が可能となる。また、新しく作られた専門用語に対し て、「用語」とう種別を付与しておくことにより、後に 新しく作られた専門用語を容易に収集して単語辞書やシ ソーラスに反映させることが容易に実現可能となる。

[0079]

【発明の効果】本発明によれば、付与されている分類情報と文書中の記載箇所との対応関係を検索時に参照できるので、分類情報を検索条件として使用することによって検索された文書の各々が作業者の必要とする文書であるか否かを容易に判定できる。

【0080】また、分類情報と文書中の記載箇所との対応関係を分類知識保守に活用することにより、高品質の

【図2】

図 2

分類情報核正部

分類コード 終正 分類知識を維持でき、作業者が指定した記載箇所に対応 する高精度な分類コード候補を認定・提示できるので、 分類作業の効率が向上する。

【図面の簡単な説明】

【図1】本実施例の概要を示す図である。

【図2】本実施例におけるシステムのブロック構成を示す図である。

【図3】本実施例におけるシステムのハードウェア構成 を示す図である。

【図4】本実施例における分類作業時の画面インタフェースの一例を示す図である。

【図5】本実施例における分類付与時の画面インタフェースの一例を示す図である。

【図6】本実施例における検索作業時の画面インタフェースの一例を示す図である。

【図7】本実施例における文書データベース内の分類情報の構成の一例を示す図である。

【図8】本実施例における分類知識の構成の一例を示す 図である。

【図9】本実施例における分類コード推定部の処理アルゴリズムを示す図である。

【図10】本実施例における分類知識更新部の処理アルゴリズムを示す図である。

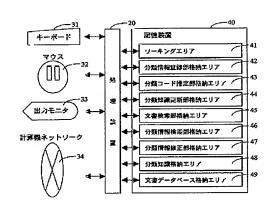
【図11】本実施例における分類コード修正時の画面インタフェースの一例を示す図である。

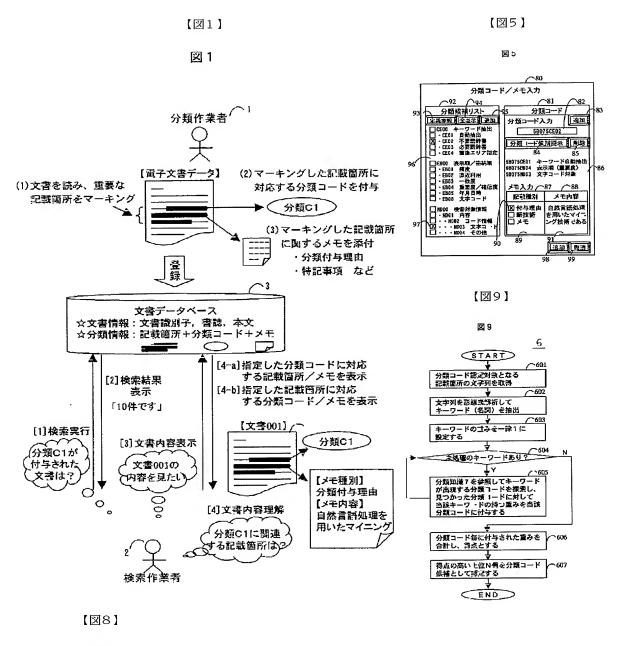
【符号の説明】

1…分類作業者、2…検索作業者、3…文書データベース、4…入出力部、5…分類情報登録部、6…分類コード推定部、7…分類知識、8…分類知識更新部、9…文書検索部、10…分類情報検索部、11…分類情報修正部

【図3】

図3





	以 8				
<u>701</u>	702	703			
キーワード	分類コード	重み			
分析	58075CE01	0.25			
分析	58075CE02	0.30			
分析	58075MD03	0.45			
マイニング	58075AA01	0.60			
マイニング	SB075CED1	0.40			
日本語	58075CE02	0.70			
日本語	58075ND03	0.30			

【図4】 【図6】 図 6 図4 **∠** 50 (70 (71 (72) 接边 集存 [終了 ۱63 106 101 102 103 104 105 特願平12-99999 前マース 東マース 北京ホ 分詞コード メモ 付与された分類 特願平12-899999 コマーク 次マーク 非並示 分類コード メモ 検案条件の指定 分類 ハード入力 追加 キーワードによる検索/ 58075CE01 キーワード自動抽出 58075ND03 文字コード対象 自然言語処理 58075CEO1 キ - ワ - ド自動抽出 58075ND03 文字コ - ド対章 自然言語処理/107 58075CE02 を用いたマイニ ング技術である。 を用いたマイニ ング技術である。 分類コードによる検索 / 114 C115 【字類名】 明練書 【発明の名称】データ分析支援方法および装置 【発明の名称】データ分析支援方法および装置 【特許請求の範囲】 【特許請求の範囲】 Л 【簡求項1】 【請求項1】 (116 (法) (116 利用者が入力装置を介して指定した文章回性に ついて、前配電子文章を再解析して当該文章属性 利用者が入力装置を介して指定した文章属性に ついて、前記電子文章を再解析して当該文章属性 ついて、町配海ナン泉を特別のして日底大学場合は データを更新し、 新定理新された文学の と更新されなかった他の属性データとの画に港在 する護集を抽扱し、前記施出された関係を記記出 力を選集を抽扱して利用者に報知することを特徴とす るデータ分析支援方法。 ついて、別記電子文章を将所付して当院文章構造 データを更新し、前配更新された文章域性データ と要新されなかった他の属性データとの間に潜在 する関係を抽出し、前配抽出された関係を前配出 力装置を介して利用者に報知することを特致とす 検定で使われた分類 区 58075CE01 自動抽出 □ 58075E804 表示順 尼戴第五录示 るデータ分析支援方法。 【発明の詳細な説明】 【発明の詳細な説明】 【発明の属する技術分野】 本発明は、デーシ分析技術に関する。 検索件数 4件 表示 【発明の属する技術分野】 本発明は、データ分析技術に関する。 □ 1. 出版平12-999999 □ 2. 出版平12-080888 □ 3. 出版平12-777777 □ 4. 出版平12-666566 109 本発明は、文章解析技術と治出された文章属性 情報を分析してそれらの間に着在する因果層係を 抽出するデータ分析技術と過去している。 分類コード沙正 าร์ก 【図7】 【図10】 【図11】 図11 図10 ⊠ 7 8 ~305/~306/~307 韓國 上位 F位 己钱 分類 分類 **~302** ~ 308 記載 記載 図画 開始 終了 記載 位置 位置 位置 メモ情報 文書 12 BE START > <u>~310</u> 分類コードの修正 種別 **~801** 内容 分類知識更新対象となる配載箇所 <u>~1101</u> の文字列および分類コードの組をすべて取得 分類コード 1103 分類コード入力 違加 001 分額 878 922 58075 CE02 <u> ~802</u> 58075CE02 データマイニングと テキストマイニング の組合せである。 1923 1088 58075 CE01 001 分類 TT001 記載策略の文字様すべてを形態を 分別□ - 下屋補提未 ■ 用以数 解析してキーワード(名詞)抽出 1104 1105 58075(28) キーワード自動抽出 58075(28) 共一・ドラ 58075(28) 大学コード分 文書全体の分類ユード 1107 58075(28) 1107 □ EB00 養示順/格納順
□ EB01 類版
□ EB01 面近別用
□ EB03 一致度
□ EB04 二素度/確保度
□ EB05 年月日時
□ EB05 年月日時 キーワードの重みを一律1に 設定する 2701 2818 --28001 AAD2 002 分類 分類コード別にキーワードをまと め、同一キーワードの重みを合計 002 新技術 F001 20002 AA81 採距離を西部的に伸 -1109 1108 1870 18151 ぱす「ルフクラブの 楽材 □ 版00 快売対象情報 □ ・N001 内容 □ ・N002 コード情報 ○ ・・ 8003 文字コー □ ・・ 8004 その他 **~805** 58075CEO1 キーワード自動抽出 58075EB04 表示機(コ要皮) 58075EB08 文字コード 58075XD03 文字コード対象 キーワードでソートした後、ある キーワードがある分類コードにお いて出現する相対的割合を算出し その値を、その分類コードにおけ るそのキーワードの重みとする 003 分類 286 301 30011 CALE

> 生成された分類知識を新しい分類 知識としてデータを置き換える (FND

フロントページの続き

004 メモ

(72)発明者 徳永 幹彦

12

35

神奈川県横浜市戸塚区戸塚町5030番地 株 式会社日立製作所ソフトウェア事業部内

HM002 再検討が必要

58023 EZ20

Fターム(参考) 5B075 ND03 NK25 NR12 PQ02 QP01 UU05 5B082 EA07 EA08 GC04

BE DE